

Cassava Quality Classification for Tapioca Flour Ingredients by Using ID3 Algorithm

Yohanssen Pratama¹, Hadi Sutanto Saragi²

¹Faculty of Informatics and Electrical Engineering, Del Institute of Technology, Sumatera Utara, Indonesia

²Faculty of Industrial Engineering, Del Institute of Technology, Jl. Sisingamangaraja, Sitoluama, Toba Samosir, Sumatera Utara, Indonesia

Article Info

Article history:

Received Jan 6, 2018

Revised Feb 15, 2018

Accepted Feb 23, 2018

Keywords:

Algorithm

Cassava

ID3

Image

Processing

ABSTRACT

Cassava is one of the main foods consumed by Indonesian people and main ingredients to make tapioca flour. In North Sumatera there is factory that produced tapioca flour to fulfill consumer demand. To be able to meet the needs of consumers and seize market share, the product must have a good quality. Product specifications are a reference for product quality and measured with 7 parameters. The seven parameters include whiteness, moisture content, spotness, ash content, thinness, residual screen, pH flour, which meets the Indonesian National Standard. In this research we use two parameters (whiteness and spotness) to determine the quality of cassava and help the factory to maintain their product quality. In here we use blob and edge detection method in image processing to detect spot and after that classified the cassava by using an ID3 algorithm.

Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Yohanssen Pratama,

Faculty of Industrial Engineering,

Del Institute of Technology,

Jl. Sisingamangaraja, Sitoluama, Toba Samosir, Sumatera Utara, Indonesia, Tel: +62 632 331234

Email: yohanssen.pratama@del.ac.id

1. INTRODUCTION

Product quality is a measure of the number of consumers who consume them. The company that produced cassava must be able to compete in market share. To attract the consumers, the company must improve the quality of the flour in production. Specification of this tapioca flour which become indicator quality of tapioca flour is important. The specification of tapioca flour is a benchmark of flour quality measured by several parameters. Each parameter must comply with the SNI.

To be able to compete in market share, especially in North Sumatera region tapioca flour must have high quality. Tapioca flour produced by the company is distributed to consumers (as foodstuff) and to industry (paper mills). The specification of the benchmark of flour quality is measured by 7 parameters. The seven parameters are whiteness, moisture content, spotness, ash content, thinness, residual screen, pH flour. Tapioca flour has 3 grades, namely Grade A, Grade B, and Grade C. However, the company only produces tapioca flour with Grade A, because about 40% of the flour that has been produced is distributed to the industry and 60% is distributed to consumers (used for foodstuff).

For the purpose to help the company to get Tapioca flour that has Grade A we decide to perform an image processing into the cassava image. The image was taken by using a camera and then preprocessed by some image processing algorithms. After going through the preprocessing stage then the image were processed by blob and edge detection method to detect spot and determine the whiteness. After the spot and whiteness data already collected then we use the ID3 algorithm to determine whether the cassava is Grade A or not.



Figure 1. Input and Output Image

Figure 1 show the image processing process that happened in the preprocessed stage. The Input image is the image that already captured by the camera. This image processed by the algorithm to produce the output image. The detail about method that used in this research will be brief explained in the next section.

2. RESEARCH METHOD

In this section will be described the methods that used to collect and process the data. Below is the method that used to get the data in the form of image.

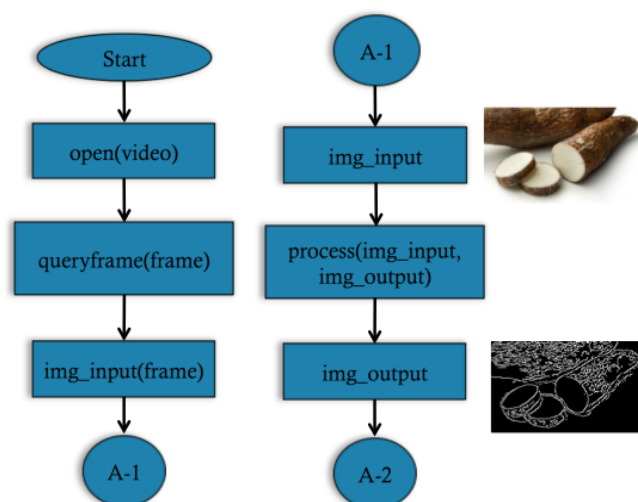


Figure 2. Method in Stage One and Two

In Figure 2 we can see the first step that need to be done is opening the video stream and take the picture from the video. The taken picture will become an input image. In this application development, we use a function in OpenCV and CvBlob library as a tool. The function that used to open the video is open function and to take picture from video we use queryframe function. Both function are member from VideoCapture class in OpenCV. After we get the picture then we process the image by using the canny edge detection algorithm. We use canny edge detection in here because the algorithm is considered better than sobel and prewitt algorithm [1]. After we processed the image with canny algorithm then the image output will be a binary image. Foreground (cassava) is represented by white color, while background (another object instead cassava) represented by black color.

Until now we already passed 2 stage, beginning from stage to get the input image from video, then continued with input image processing with canny algorithm [1]. With 2 stages above we can already distinguish between the cassava and another object that appears on the frame. But to distinguish between a good cassava and a bad one we must be able to detect spots on cassava. The cassava image's that has been detected as foreground image will be separated after all. So now we can focus on the spot that exist in the cassava.

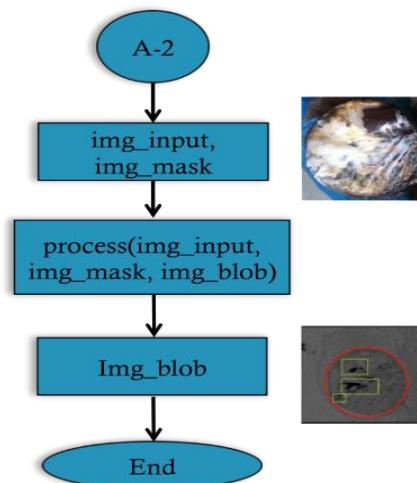


Figure 3. Method in Stage Three

The spots that exist in the cassava now will become foreground object and another region in the cross section of cassava will be a background. The spots that detected will be represented as blob first, then after we could detect a blob, we could use the number of blob information to determine quality of cassava. The input image that use in the stage 3 is a grayscale image from a cassava cross section and the image which is already masked. Masked perform in here to reduce the noise and we used a Gaussian blur to perform masking.

Table 1. Product Specification

No	Parameter	Specification
1	Whiteness	$\geq 91.0 \%$
2	Thinness	$\geq 99.50 \%$
3	Spots	≤ 5
4	moisture content	12.00-13.00 %
5	ash content	0.02-0.06 %
6	Residual screen	325 (max 0.05 %),
7	pH flour	5.6 - 6.5

If we see Table 1, we found that the specification that according to consumer expectations is if the number of spots in cassava is smaller than 5 spots. From here we could characterized the cassava quality based on number of spots that exist. Classification was made by the condition in the Table 2 (there are 3 kind condition). If the number of spots smaller than 3, then we categorize cassava into good quality and other than that the cassava will categorize into medium or bad quality.

Table 2. Cassava Classification Based on Spots

Quality/Grade	Number of spots
Good (Grade A)	Spots < 3
Medium (Grade B)	$3 \leq \text{Spots} \leq 5$
Bad (Grade C)	Spots > 5

Besides using the number of spot to determine the cassava quality, we use whiteness as parameter that determine the quality.

Table 3. Cassava Classification Based on Whiteness

Quality/Grade	Whiteness
Good (A)	$\geq 91\%$
Medium (B)	89 %
Bad (C)	$\leq 89\%$

Cassava Grade or Quality that produced by the company is grade A cassava. So we use an image processing again to determine the whiteness of cassava. To get the whiteness percentage, we compare the pixel value of an input image with the ideal value for the whiteness in 8-bits color system (255,255,255). Then we averaging spectral distance from each pixel value if the difference is smaller than 9% then we could categorize the cassava into grade A (based on Table 3). Below is the spectral distance formula that used:

$$D = \sqrt{\sum_{i=1}^n (d_i - e_i)^2}$$

i is a band (dimension) and d_i is value of pixel in band i , otherwise e_i is value of pixel e in band i . For our case d has value 255 and e value based on the pixel intensity in input image. The input image that we used to get this whiteness score is on grayscale color space.

Until now we already have 2 parameters to make classification, the last parameter that we used is moisture content in the cassava. To get moisture content we use moisture meter (Figure 4).



Figure 3. Moisture Meter

For moisture we only have 2 discrete parameter values, there are good and bad. Good if the moisture content is around 12-13%, other than that will be classified into bad category. After we get 3 parameters that needed to make a classification (number of spots, whiteness, and moisture) then we use them as input data for learning algorithm. Below is the example of collected data (Table 4).

Table 4. Data Example

Spot	Whiteness	Moisture	Quality
Medium	Good	Good	Good
Bad	Medium	Bad	Bad
Medium	Good	Bad	Good
Medium	Bad	Good	Bad
Good	Medium	Bad	Bad

The data above is already past the preprocessing stage by using discretization method (to discretize the data we used the rule from Table 2 and Table 3). In this research we used a decision tree learning method to process the data because the problem that would be solved have a target function which has a discrete output and there are error and data loss possibility from the training data. The decision tree algorithm that used in here is ID3 algorithm. ID3 algorithm use statistic properties that referred as information gain and this information gain will be used to select a candidate to build the tree [2]. To get the information gain accurately

we need the entropy first. Entropy serves to measure the amount of information that contained in an attribute. Below is the entropy formulation:

$$Entropy(S) = - \sum_{i=1}^c p_i \log_2 p_i$$

If the target attribute has different value as much as c (in our case the target is Quality that only have 2 attribut: Good or Bad) and p_i is part of S that belong to c . After we get the entropy we could get the attribute effectiveness value in classifying a training data or we usually called it as information gain [2]. The information gain, $Gain(S, A)$ with the set of Examples S towards attribute A can be defined as follows:

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Where $Values(A)$ is a set of possible values in attribute A , and S_v is a subset of S where attribute A has a value of v (e.g. $S_v = \{s \in S \mid A(s) = v\}$). Whereas $|S_v|$ is the number of elements in S_v and $|S|$ is the number of elements in S . The first part of the gain equation is entropy of S , while the second part is the expected value of entropy S after being partitioned using attribute A .

3. RESULTS AND ANALYSIS

The result that we get from the stage 3 of data acquisition is number of spot. First we tried to capture the image from the sample that has a good quality and then we will get the result as shown in Figure 5.

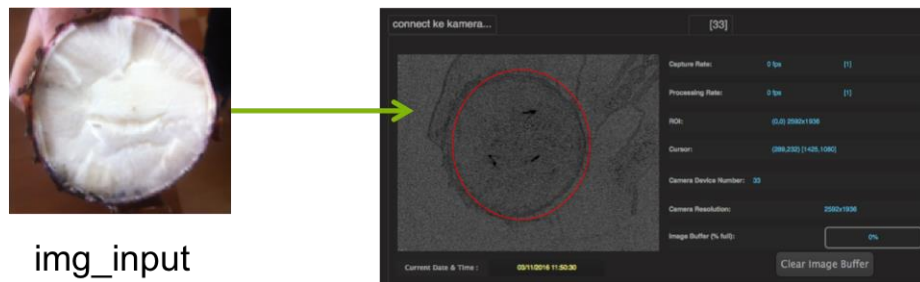


Figure 5. Spot Detection in Good Cassava

There is no blob detected in Figure 5, although there was a small black spot that detected. The small spot not detected here because we already determine the minimum and maximum area of blob which can be categorized as spot. If the spot is too small, we not categorized it as a spot. The minimum and maximum size area of blob is decided by applying a learning process. After that we also capture the image from the bad quality sample (Figure 6).



Figure 6. Spot Detection in Bad Cassava

In Figure 6, there are 3 blobs detected and pass the minimum area size that already decided. So from here we can conclude that the good cassava has zero spot and on the other hand the bad cassava has three numbers of spots. There are three parameters that we used to determine whether the cassava is bad or good. So we still need the information about whiteness and moisture. In bad cassava we find that the whiteness is below 89% and moisture is more than 13%, so after we get the three parameters information we could decide that the cassava quality is bad. In Table 4 we could see some data sample that we've already collected. For testing the accuracy is used dataset as follows:

Table 5. Dataset

Datasets Number	Training Data (60%)	Test Data (40%)
400	240	160

So if we see Table 5, we use 60% from all data sample as training data and the rest will be a test data. The Figure 7 below is an illustration of the accuracy testing process that performed:

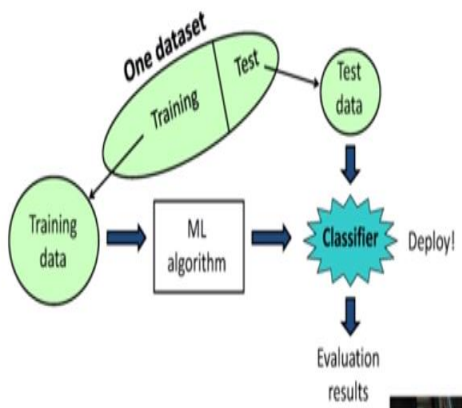
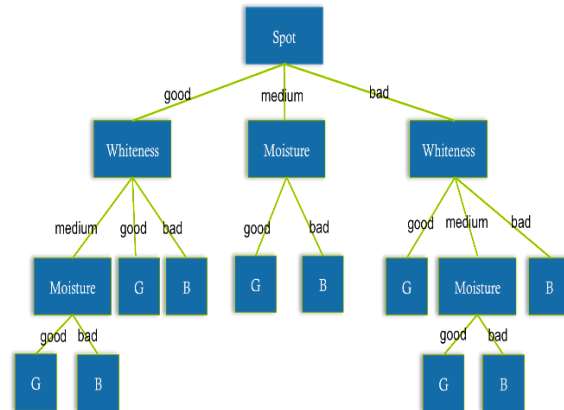
Figure 7. Accuracy testing process
(weka.waikato.ac.nz)

Figure 8. Decision Tree

The training data will be use to train the ID3 algorithm so it will have it's own classifier in tree form. The rule in the classifier (tree) will be used to classifying the test data. After that we could get the accuracy from the evaluation results. Below is the decision tree that obtained after we process the training data (Figure 8):

The accuracy that we get from the decision tree above classifier is 84,7328. We also try to compare the performance of the id3 algorithm with 8 other algorithms. The accuracy of the 8 algorithms is determined by using weka and below is the result for making some comparison:

Table 6. Accuracy Comparison

No.	Algorithm Name	Accuracy (%)
1.	Bayesnetwork	77.8626
2.	Naivebayes	77.8626
3.	RandomTree	79.3893
4.	ID3	84.7328
5.	K*	80.1527
6.	RandomForest	83.9695
7.	K Nearest Neighbour	84.7328
8.	FunctionalTrees	85.4962
9.	J48	87.0229

From Table 6 we can see that ID3 algorithm is not the best algorithm but it still have a good performance if we compare with another algorithm.

4. CONCLUSION

From our research, we can conclude that to determine a good quality cassava we could use image processing to collect the data and use an id3 algorithm to processed the data and produce some decision. From three parameters that used here, there are whiteness, moisture, and number of spot we could get 84.7328% accuracy to classify the cassava based on it quality. For futures work we could add more parameters to increase the accuracy and use another classifier algorithm such as neural network which more promising when doing classification.

REFERENCES

- [1] Canny J. A Computational Approach to Edge Detection. *Journal IEE-PAMI*. 1986; 8.(6): 679-698.
- [2] Mitchell M. Machine Learning. Edition. McGraw-Hill. 1997: 55-60.
- [3] Haijian S. Best-First decision tree learning. Master's Thesis. Hamilton: Postgraduate University of Waikato; 2007.
- [4] Narendra V and Hareesh K. Quality Inspection and Grading of Agricultural and Food prodcuts by Computer Vision-A Review. *International Journal of Computer Applications (0975-8887)*. 2(1).
- [5] Yam K and Spyridon E. A Simple Digital Imaging Method for Measuring and Analyzing Colour of Food Surfaces. *Journal of Food Engineering*. 2003; 61:137-142
- [6] Kodagali A and Balaji S. Computer Vision and Image Analysis based Techniques for Automatic Characterization of Fruits- a Review. *International Journal of Computer Applications*. 2012; 50(6).
- [7] Lpez-Garca F, Andreu-Garca G, Blasco J, Aleixos N, and Valiente M. Automatic detection of skin defects in citrus fruits using a multivariate image. *Computers and electronics in Agriculture*. 2010; 71:189-197.
- [8] Timmermans A. Computer Vision System for Online Sorting of Pot Plants Based on Learning Techniques. *ActaHorticulturae*. 1998; 421:91-98.
- [9] Sardar H. Quality Analysis in grayscale color using visual appearance of guava fruit. *International Journal of Engineering Sciences*. 2013; 46-56.
- [10] Rocha A, Hauagge D, Wainer J, Goldenstein S. Automatic fruit and vegetable classification from images. *Computers and Electronics in Agriculture*. 2010; 70:96-104.
- [11] Yousef A. Computer vision based date fruit grading system: Design and implementation. *J of King Saud University. Computer and Information Sciences*; 2011:23:29-36.
- [12] Seng W and Mirisaee S. A new method for fruits recognition system. *International Conference on Electrical Engineering and Informatics*. 2009: 130-134.
- [13] Dadwal M and Banga V. Estimate Ripeness Level of fruits using RGB Color Space and Fuzzy Logic Technique. *International Journal of Engineering and Advanced Technology*. 2012; 2(1): 225-229.
- [14] Benhura C, Albert M, Muchuweti M, Gombiro E. Assesment of the Colour of Parinari Curatellifolia Fruit using an image processing computer software package. *International Journal of Agricultural and Food Research*. 2013; 2(4): 41-48.
- [15] Yudong Z and Lenan W. Classification of fruits using computer vision and multiclass support vector machine. *Sensors*. 2012; 12: 12489-12505